IT Unit 3

Topic 2

**Identifying patterns and relationships between data**, 152-165

1. What is the purpose of using statistics when working with large data sets?

*Convert large quantities of raw data into small, informative, meaningful summaries. Allow patterns & relationships to be identified.*

1. List the 3 most useful statistical concepts from table 3.7, p 152.

* Average: The single value that gives the most representative summary value of a range of numbers; mean, mode and median are different ways of averaging data under different circumstances.
* Standard Deviation: Indicates how consistent the values are in a data set. Are the values all close to the average, or do they vary greatly? A low standard deviation means that the mean is truly representative of all the data and can be trusted. A high standard deviation suggests you should not trust the mean.
* Correlation: Do two data sets show the same trends, as if they might be connected in some way?

**Basic statistics**

1. Distinguish between the three methods of calculating an average.

The mean (Central value; Add up all numbers and divide by how many numbers are in the set), median (Middle number; put numbers is numerical order and chose the number that is halfway through the set), and mode (most frequent appearing number).

1. What is the purpose of the standard deviation?

To measure how much a data set varies from the mean of the data.

**Correlation and causality**, p 156

1. Why do you need to be careful when looking at patterns in data in terms of cause and effect?

Your brain may see two sets of data that are completely unrelated and view it as a pattern where the brain may then see a misleading pattern that should not exist.

**Data visualisations, p 157**

1. What are data visualisations used for?

Data visualisations are used for creating (analogue) lines, shapes and colours to represent (digital) numbers to make the data easy to understand and interpret.

**Queries and searches,** p 159

1. How can queries, searches, filtering and sorting be used when analysing large data sets?

Queries and searches will help you to locate the patterns in your data sets so that you can find the larger story that your outcome data is telling you.

**Conditional formatting,** p 164

1. Explain the process of conditional formatting.

Conditional formatting is changing the appearance of data automatically based on its current value in spreadsheets and databases. This highlights data that may be of special interest, also possibly highlight errors inside the database. (p164 example process).